Scorecard design for fraud detection
(with text mining, predictive modelling
and social network theory)

Terisa Roberts
Philip Pretorius

28 August 2009
State of the Global Insurance Industry

- Global recession
- Acquisitions and Mergers
- Stronger competition, especially from non-traditional insurance companies
- Tighter regulation

“Fraudulent claims have doubled in the first three months of 2009” - Allianz Insurance, United Kingdom
State of South African Insurance Industry

- Emerging market
- Cost of insurance more expensive in relation with people’s income
- South African insurance industry has tendency to follow international trends
Insurance Fraud in South Africa

- 5%-10% of claims said to be fraudulent*
- Costing insurance industry R2 billion per year
- Set up of the South African Insurance Crime Bureau in 2008

“Between 8% and 35% of short-term insurance claims paid out to policyholders annually are fraudulent.” – Insurance Companies, South Africa

*South African Insurance Association, 2008
Investigated Cases

- Large proportion unfounded
- Mean (investigation period) = 80 days
- Other = internal fraud etc.
Challenges in Fraud Management

• Fraud detection reactive, rather than pro-active
• Diagnostic indicators commonly used – but not tested
• Special Investigations Unit – limited resources, fraud management spreadsheets
• “Feedback loop” not complete
• Infrequent event data, “tip of the iceberg”
• Detection techniques used with varying degrees of success
  – Redundant complexity (out-dated rules)
  – Disparate systems and utilization
  – Deteriorating performance over time
Key Properties* of a Suspicious Activity Assessment System

- Accurate
- Fast
- Cost-effective
- Flexible
- Consistent
- Reliable
- Easy to interpret
- Adaptive

*Abrahams, 2008
Fraud Detection Techniques

- Diagnostic Fraud Indicators
- 3rd party data searching
- Anomaly Detection
- Profiling
- Supervised/Unsupervised Methods
- Artificial Intelligence
- Text Mining
- Social Network Theory
Artificial Intelligence

- Machine Learning
- Neural Networks
- Expert Systems

- Complex
- Sensitive to noise
- Interpretability challenges
Text Mining

• Natural Language Processing
• Semantics – meaning of words
• Syntax – structural relationship between words
• Text Parsing
• Dimension Reduction Techniques
Social Network Theory

*Fast unfolding of communities in large networks, 2008*
### (Hypothetical) Fraud Risk Scorecards

#### Probability of Fraud

| Claim level - Quantitative factors | Claim amount out of normal bounds for loss class | Vehicle burnt / total theft with coverage recently increased | Dubious location of loss | Recent similar claim | No towing charges, although extensive damage | Claim level - Qualitative factors | Lack of witnesses | Attitude: Aggressive/Evasive/Vague | Threaten to obtain attorneys | Policy Information | Claim within 3 months of inception | Recent cover increase | Customer Information | New customer | Insured moved to lower income risk address | Mobile phone contact only | Occupation |
|-----------------------------------|-------------------------------------------------|------------------------------------------------------------|-------------------------|----------------------|---------------------------------------------|---------------------------------|-----------------|-------------------------------|-----------------|-----------------|---------------------------|-----------------|-----------------|------------------------|-----------------------|---------------------|
| Fraud bureau scores | Credit Bureau scores | Social Networks | Claimant’s attorney syndicate | Suspicious home address | Suspicious broker |

#### Potential Loss

| Claim level - Quantitative Factors | Hijack / Burnt out vehicle | Insured verified coverage just prior to loss date | Claim level - Qualitative Factors | Information inconsistencies | Policy information | High premium payments compared to verifiable legitimate income | Repeated and unexplained change of beneficiary | Unusually high commission paid to broker/intermediary | Total sum insured | Customer Information | Geographic region of home address | Temporary post office box | Insured recently divorced | Fraud Bureau Score | Credit Bureau Score | Social Networks | Suspicious home address |

#### Likelihood

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>S</th>
<th>M</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost certain</td>
<td>Minor</td>
<td>Moderate</td>
<td>Major</td>
</tr>
<tr>
<td>Likely</td>
<td>Minor</td>
<td>Moderate</td>
<td>Major</td>
</tr>
<tr>
<td>Possible</td>
<td>Minor</td>
<td>Moderate</td>
<td>Major</td>
</tr>
<tr>
<td>Unlikely</td>
<td>Minor</td>
<td>Moderate</td>
<td>Major</td>
</tr>
<tr>
<td>Rare</td>
<td>Minor</td>
<td>Moderate</td>
<td>Major</td>
</tr>
</tbody>
</table>

#### Consequences

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>S</th>
<th>M</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almost certain</td>
<td>Insignificant</td>
<td>Minor</td>
<td>Moderate</td>
</tr>
<tr>
<td>Likely</td>
<td>Insignificant</td>
<td>Minor</td>
<td>Moderate</td>
</tr>
<tr>
<td>Possible</td>
<td>Insignificant</td>
<td>Minor</td>
<td>Moderate</td>
</tr>
<tr>
<td>Unlikely</td>
<td>Insignificant</td>
<td>Minor</td>
<td>Moderate</td>
</tr>
<tr>
<td>Rare</td>
<td>Insignificant</td>
<td>Minor</td>
<td>Moderate</td>
</tr>
</tbody>
</table>
Conclusion

• Fraud remains a big challenge
• A pro-active and accurate suspicious activity assessment system should enable insurance companies to
  – Prioritise and improve quality and quantity of investigations
  – Reduce fraud expenditure
  – Uncover organised crime
• By utilising volumes of internal, external, structured and unstructured data
• Whilst maintaining an easy to implement and easy to interpret design
Some References

• Abrahams C., Zhang M.; “Credit Risk Assessment The New Lending System for Borrowers, Lenders and Investors”; 2009; p1, p290-291
• Blondel V., Guillaume J., Lambiotte R., Lefebvre E.; “Fast unfolding of communities in large networks”; 2008
• Bolton R., Hand D.; “Statistical Fraud Detection Review”; 2002
• Caudill S., Ayuso M., Guill´en M.; “Fraud Detection using a multinomial logit model with missing information”; 2005
• Lilley P.; Dirty Dealing – The untold truth about global money laundering, international crime and terrorism”; 2003; p7, p230,p225
• Morley N.,Ball L., Ormerod T.; “How the detection of insurance fraud succeeds and fails”; 2006
• O’Gara J.D.; “Corporate Fraud – Case Studies”; 2004; p131-148, p155-157
• Phua C., Lee V., Smith K., Gayler R.; “A Comprehensive Survey of Data Mining-based Fraud Detection Research”; 2002
• Reuter P., Truman E., “Chasing Dirty Money”; 2004; p29, p33
• Rowe R., Creamer G., Hershkop S., Stolfo S.; “Automated Social Hierarchy Detection through Email Network Analysis”
Thank you